

UCLA STAT 13
**Introduction to Statistical Methods for
the Life and Health Sciences**

● **Instructor:** Ivo Dinov,
Asst. Prof. In Statistics and Neurology

● **Teaching Assistants:**
Ming Zheng, Annie Che
UCLA Statistics

University of California, Los Angeles, Winter 2004
http://www.stat.ucla.edu/~dinov/courses_students.html

STAT 13, UCLA, Ivo Dinov Slide 1

Chapter 6: Continuous Random Variables

- Continuous Random Variables
- The Normal Distribution
- Sums and differences of random quantities

STAT 13, UCLA, Ivo Dinov Slide 2

Dietary intake of carbohydrate (mg/day) for N=5929 people from a variety of work environments. Standardized histogram plot is unimodal but skewed to the right (high values). Vertical scale is (relative freq.)/(interval width) = $f_j/(N*m)$. The proportion of the data in [a : b] is the area under the standardized histogram on the range [a : b].

(a) Standardized histogram

(b) Area between a = 225 and b = 375 shaded

Slide 3 STAT 13, UCLA, Ivo Dinov

Superposition of a smooth curve (density function) on the standardized histogram (left panel). Area under the density curve on [a : b] = [225 : 375] is analytically computed to be: 0.486 (right panel), which is close to the empirically obtained estimate of the area under the histogram on the same interval: 0.483 (left panel).

(c) With approximating curve

(d) Area between a = 225 and b = 375 shaded

Slide 4 STAT 13, UCLA, Ivo Dinov

Standardized histograms

For a standardized histogram:

- The vertical scale is $Relative_frequency / Interval_width$
- Total area under histogram = 1
- Proportion of the data between a and b is the area under histogram between a and b

Slide 5 STAT 13, UCLA, Ivo Dinov

Probability and areas

For a continuous X

- the probability a random observation falls between a and b = area under the density curve between a and b.

Slide 6 STAT 13, UCLA, Ivo Dinov

Sampling from the distribution in Fig. 6.1.1

(a) Dot plots of 6 sets of 15 random observations

Note the fair amount of **intra- and inter-group variability**. What does that mean? Is that **normal or expected**?

Slide 7 STAT 13, UCLA, Ivo Dinov

(b) Histograms with density curve superimposed

Slide 9 STAT 13, UCLA, Ivo Dinov

Visualizing the population mean

The **population mean** is the imaginary value of X where the **density curve balances**

Slide 10 STAT 13, UCLA, Ivo Dinov

Interval endpoints and continuous variables

Recall a **continuous variable** is one where the domain has no gaps in between the values the variable can take.

In calculations involving a **continuous random variable** we *do not have to worry* about whether interval endpoints are included or excluded.

Slide 11 STAT 13, UCLA, Ivo Dinov

Review

- How does a **standardized histogram** differ from a **relative-frequency histogram**? **raw histogram**? (f_j/mn)
- What graphic feature conveys the **proportion of the data** falling into a class interval for a **standardized histogram**? for a **relative-frequency histogram**?
(area=width · height = $m f_j/mn = f_j/n$)
- What are the **two fundamental ways** in which **random observations arise**? (Natural phenomena, sampling experiments – choose a student at random and use the lottery method to record characteristics, scientific experiments - blood pressure measure)
- How does a **density curve** describe probabilities?
(The probability that a random obs. falls in $[a;b]$ is the area under the PDF on the same interval.)

Slide 12 STAT 13, UCLA, Ivo Dinov

Review

- What is the **total area** under both a **standardized histogram** and a **probability density curve**? (1)
- When can **histograms** of data from a random process be relied on to **closely resemble the density curve** for that process? (large sample size, small histogram bin-size)
- What characteristic of the **density curve** does the **mean** correspond to? (imaginary value of X , where the **density curve balances**)

Slide 13 STAT 13, UCLA, Ivo Dinov

Review

- Does it matter whether **interval endpoints** are **included** or **excluded** when we calculate probabilities for a **continuous random variable** from the area? (No)
- Why? (Area[a:b] = Area(a:b))
- Are **discrete variables** the **same** or **different** in this regard, interval endpoint not effecting the area? (Different)

Slide 14 STAT 13, UCLA, Ivo Dinov

Two standardized histograms with approximating Normal density curve

(a) Chest measurements of Quetelet's Scottish soldiers (in.)

(b) Heights of the 4294 men in the workforce database (cm)

Slide 15 STAT 13, UCLA, Ivo Dinov

The Normal distribution density curve

- Is symmetric about the mean! Bell-shaped and unimodal.
- Mean = Median!

$N(\mu, \sigma)$

50% 50%

Mean μ

Slide 16 STAT 13, UCLA, Ivo Dinov

Effects of μ and σ

(a) Changing μ

shifts the curve along the axis

$\sigma_1 = \sigma_2 = 6$

$\mu_1 = 160$ $\mu_2 = 174$

Mean is a measure of ...

central tendency

(b) Increasing σ

increases the spread and flattens the curve

$\sigma_1 = 6$

$\sigma_2 = 12$

$\mu_1 = \mu_2 = 170$

Standard deviation is a measure of ...

variability/spread

Slide 17 STAT 13, UCLA, Ivo Dinov

Understanding the standard deviation: σ

Probabilities/areas and numbers of standard deviations for the Normal distribution

Shaded area = 0.683	Shaded area = 0.954	Shaded area = 0.997
$\mu - \sigma$ $\mu + \sigma$	$\mu - 2\sigma$ $\mu + 2\sigma$	$\mu - 3\sigma$ $\mu + 3\sigma$
68% chance of falling between $\mu - \sigma$ and $\mu + \sigma$	95% chance of falling between $\mu - 2\sigma$ and $\mu + 2\sigma$	99.7% chance of falling between $\mu - 3\sigma$ and $\mu + 3\sigma$

Slide 18 STAT 13, UCLA, Ivo Dinov

Probabilities supplied by computer programs – Cumulative (lower-tail) probabilities

Area = $\text{pr}(X \leq x)$

Area = $\text{pr}(X \leq x)$


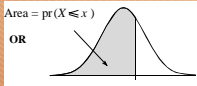
OR

Areas in [0:Z] of the Std.Normal Distribution Are obtained by STATA command `ztable`

Slide 19 STAT 13, UCLA, Ivo Dinov

Probabilities supplied by computer programs – Cumulative (lower-tail) probabilities

Problem: To find $P(X \leq 180)$, when $\mu=174$ and $\sigma=6.57$
Convert to Standard units: $Y = (X - \mu) / \sigma = 6/6.57 = 0.91$
Look-up the Normal Distribution Table: 0.3186
Final cumulative (lower-tail) result: $0.5 + 0.3186 = 0.819$

Slide 20 STAT 13, UCLA, Ivo Dinov

Finding Help with STATA

• http://www.stat.ucla.edu/~dinov/courses_students.dir/STAT13_Fall01/STAT13_labs.html

Slide 21 STAT 13, UCLA, Ivo Dinov

Basic method for obtaining probabilities

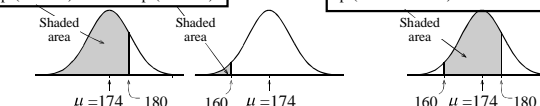
- Sketch a **Normal curve**, marking the mean and other values of interest.
- **Shade the area** under the curve that gives the desired probability.
- Devise a way of getting the desired area from **lower-tail areas**.
- Obtain component lower-tail **probabilities from a computer program**

Slide 22 STAT 13, UCLA, Ivo Dinov

(a) Computing $pr(160 < X \leq 180)$

Programs supply $pr(X \leq 180)$ and $pr(X \leq 160)$

We want $pr(160 < X \leq 180) = \text{difference}$



$pr(160 < X \leq 180) = pr(X \leq 180) - pr(X \leq 160)$

Slide 23 STAT 13, UCLA, Ivo Dinov

Fig. 6.2.4(c)

(c) More Normal probabilities
(values obtained from Minitab)

b	$pr(X \leq b)$	a	$pr(X \leq a)$	$pr(a < X \leq b) = \text{difference}$
167.6	0.165	152.4	0.001	0.164
177.8	0.718	167.6	0.165	0.553
177.8	0.718	152.4	0.001	0.717
182.9	0.912	167.6	0.165	0.747

Note: 152.4cm = 5ft, 167.6cm = 5ft 6in., 177.8cm = 5ft 10in., 182.9cm = 6ft

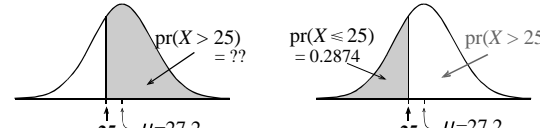
From Chance Encounters by C.J. Wild and G.A.F. Seber. © John Wiley & Sons, 2000.

Slide 24 STAT 13, UCLA, Ivo Dinov

Obtaining an upper-tail probability

We want $pr(X > 25) = ??$

Programs supply $pr(X \leq 25) = 0.2874$



Since total area under curve = 1, $pr(X > 25) = 1 - pr(X \leq 25)$

Generally, $pr(X > x) = 1 - pr(X \leq x)$

Slide 25 STAT 13, UCLA, Ivo Dinov

Review

- What features of the Normal curve do μ and σ visually correspond to? (point-of-balance; width/spread)
- What is the probability that a random observation from a normal distribution is smaller than the mean?
 - (0.5) larger than the mean? (0.5) exactly equal to the mean? (0.0) Why?

Slide 26 STAT 13, UCLA, Ivo Dinov

Review

- Approximately, what is the probability that a random observation from a normal distribution falls within 1 standard deviation (SD) of the mean? (0.68) 2 SD's? (0.95) 3 SD's? (0.997)
- Computer programs may provide **cumulative** or **partial probabilities** for the **normal distribution**. What is the difference between these? Can we get one from the other?

Slide 27 STAT 13, UCLA, Ivo Dinov

The inverse problem – Percentiles/quantiles

(a) p-Quantile

Programs supply x_p
x-value for which $pr(X \leq x_p) = p$

$x_p = ???$

(b) 80th percentile (0.8-quantile) of women's heights

Normal($\mu = 162.7$)

prob = 0.8

$\mu = 162.7$ $x_{0.8} = ???$

Program returns
Thus 80% lie below

80% of people have height below the **80th percentile**. This is EQ to saying there's **80% chance** that a random observation from the distribution will fall below the **80th percentile**.

(c) Further percentiles of women's heights										
Percent	1%	5%	10%	20%	30%	70%	80%	90%	95%	99%
Probn	0.01	0.05	0.1	0.2	0.3	0.7	0.8	0.9	0.95	0.99
Percentile (or quantile)	148.3	152.5	154.8	157.5	159.4	166.0	167.9	170.6	172.4	177.4
(in cm)	49.0"	50"	50"	52"	52"	55"	55"	57"	57"	58"

The **inverse problem** is what is the height for the 80th percentile/quantile? So far we studied given the height value what's the corresponding percentile?

Slide 28 STAT 13, UCLA, Ivo Dinov

The inverse problem – upper-tail percentiles/quantiles

Obtaining an inverse upper-tail probability

"What value gives the top 25%?"

What does this say about the lower tail?

$\mu = 162.7$ $???$

$\mu = 162.7$ $???$

Slide 29 STAT 13, UCLA, Ivo Dinov

Review

- What is meant by the 60th percentile of heights?
- What is the **difference** between a **percentile** and a **quantile**? (percentile used in expressing results in %, whereas quantiles used to express results in term of probabilities)
- The **lower quartile**, **median** and **upper quartile** of a distribution correspond to **special percentiles**. What are they? express in terms of quantiles. (25%, 50%, 75%)
- Quantiles** are sometimes called inverse **cumulative probabilities**. Why?

Slide 30 STAT 13, UCLA, Ivo Dinov

Standard Normal Curve

- The standard normal curve is described by the equation:

$$y = \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}$$

Where remember, the natural number $e \sim 2.7182\dots$
We say: **X-Normal(μ, σ)**, or simply **X-N(μ, σ)**

Slide 31 STAT 13, UCLA, Ivo Dinov

Standard Normal Approximation

- The **standard normal curve** can be used to estimate the percentage of entries in an interval for any process. Here is the protocol for this approximation:
 - Convert the interval (we need to assess the percentage of entries in) to **standard units**. We saw the algorithm already.
 - Find the corresponding area under the normal curve (from tables or online databases);

Data

Transform to Std.Units

What percentage of the density scale histogram is shown on this graph?

Compute %

Report back %

Slide 32 STAT 13, UCLA, Ivo Dinno

General Normal Curve

- The **general normal curve** is defined by:
 - Where μ is the **average** of (the symmetric) normal curve, and σ is the **standard deviation** (spread of the distribution).

$$y = \frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}}$$

- Why worry about a **standard** and **general** normal curves?
- How to convert between the two curves?

$\mu = 0.8$
 $\sigma = 1.8$

$\mu = 12$
 $\sigma = 12$

Slide 33 STAT 13, UCLA, Ivo Dinno

Areas under Standard Normal Curve – Normal Approximation

- Protocol:**
 - Convert the interval (we need to assess the percentage of entries in) to **Standard units**. Actually convert the end points in Standard units.
 - In general, the transformation $X \rightarrow (X-\mu)/\sigma$, **standardizes** the observed value X , where μ and σ are the **average** and the **standard deviation** of the distribution X is drawn from.
 - Find the corresponding area under the normal curve (from tables or online databases);
 - Sketch the normal curve and shade the area of interest
 - Separate your area into individually computable sections
 - Check the Normal Table and extract the areas of every sub-section
 - Add/compute the areas of all sub-sections to get the total area.

Slide 34

Obtaining central range for symmetric distributions

What values contain the **central 50%**?

What does that say about the **lower tails**?

0.5

$a = ??$ $b = ??$

0.25

$a = ??$ $b = ??$

0.75

$b = ??$

Obtain **b** from program

0.25

$a = ??$

Obtain **a** from program

Slide 35 STAT 13, UCLA, Ivo Dinno

The z-score

- The **z-score** of x is the number of standard deviations x is from the mean. (Body-Mass-Index, BMI)

TABLE 6.3.1 Examples of z -Scores

X	z -score = $(x - \mu) / \sigma$	Interpretation
Male BMI values (kg/m ²)		
25	$(25-27.3)/4.1 = -0.56$	25 kg/m ² is 0.56 sd's below the mean
35	$(35-27.3)/4.1 = 1.88$	35 kg/m ² is 1.88 sd's above the mean
Female heights (cm)		
155	$(155-162.7)/6.2 = -1.24$	155cm is 1.24 sd's below the mean
180	$(180-162.7)/6.2 = 2.79$	180cm is 2.79 sd's above the mean

Male BMI-values: $\mu = 27.3$, $\sigma = 4.1$ Females heights: $\mu = 162.7$, $\sigma = 6.2$

- Which ones of these are **unusually** large/small/away from the **mean**?

Slide 36 STAT 13, UCLA, Ivo Dinno

The standard Normal distribution

Standard Normal distribution:

mean(μ) = 0, SD(σ) = 1

Slide 37 STAT 13, UCLA, Ivo Dinno

Working in standard units

What values contain the **central 90%**?

What does that say about the lower tail?

Obtain z from program
[Program returns 1.6449]

- The **z -score** of x is the number of standard deviations x is away from the mean.

Slide 38 STAT 13, UCLA, Ivo Dinov

Working in standard units (# of SD's)

TABLE 6.3.2 Central Ranges

Percentage	z	Male BMI values		Female heights	
		$\mu - z\sigma$	$\mu + z\sigma$	$\mu - z\sigma$	$\mu + z\sigma$
80%	1.2816	22.05	32.55	154.8	170.6
90%	1.6449	20.56	34.04	152.5	172.9
95%	1.9600	19.26	35.34	150.5	174.9
99%	2.5758	16.74	37.86	146.7	178.7
99.9%	3.2905	13.81	40.79	142.3	183.1

Male BMI values: $\mu=27.3, \sigma=4.1$
Females heights: $\mu=162.7, \sigma=6.2$

Standardizing $Z = (X - \mu) / \sigma$
Inverting $X = Z\sigma + \mu$

Slide 40 STAT 13, UCLA, Ivo Dinov

TABLE 6.3.3 Using z -score tables

As an example, we shall find $\text{pr}(Z \leq 1.1357)$ using part of the table given in Appendix A4 (reproduced below).

Step 1: Correct the z -value to two decimal places, that is, use $z = 1.14$.

Step 2: Look down the z column until you find 1.1. This tells you which row to look in.

Step 3: The second decimal place, here 4, tells you which column to look in.

Step 4: The entry in the table corresponding to that row and column is $\text{pr}(Z \leq 1.14) = 0.873$

z	0	1	2	3	4	5	6	7	8	9
1.0	.841	.844	.846	.848	.851	.853	.855	.858	.860	.862
1.1	.864	.867	.869	.871	.873	.875	.877	.879	.881	.883
1.2	.885	.887	.889	.891	.893	.894	.896	.898	.900	.901
1.3	.903	.905	.907	.908	.910	.911	.913	.915	.916	.918
1.4	.919	.921	.922	.924	.925	.926	.928	.929	.931	.932

Slide 41 STAT 13, UCLA, Ivo Dinov

Quincunx

Slide 42 STAT 13, UCLA, Ivo Dinov

Continuous Variables and Density Curves

- There are no gaps between the values a continuous random variable can take.
- Random observations arise in two main ways: (i) by sampling populations; and (ii) by observing processes.

Slide 43 STAT 13, UCLA, Ivo Dinov

The density curve

- The probability distribution of a continuous variable is represented by a density curve.
 - Probabilities** are represented by **areas under the curve**,
 - the probability that a random observation falls between a and b equal to the area under the density curve between a and b .
 - The total area under the curve equals 1.
 - The population (or distribution) mean $\mu_X = E(X)$, is where the density curve balances.
 - When we calculate probabilities for a continuous random variable, it does not matter whether interval endpoints are included or excluded.

Slide 44 STAT 13, UCLA, Ivo Dinov

For any random variable X

- $E(aX + b) = a E(X) + b$ and $SD(aX + b) = |a| SD(X)$

Slide 45 STAT 13, UCLA, Ivo Dinov

The Normal distribution

$$X \sim \text{Normal}(\mu_x = \mu, \sigma_x = \sigma)$$

Features of the Normal density curve:

- The curve is a symmetric bell-shape centered at μ .
- The standard deviation σ governs the spread.
 - 68.3% of the probability lies within 1 standard deviation of the mean
 - 95.4% within 2 standard deviations
 - 99.7% within 3 standard deviations

Slide 46 STAT 13, UCLA, Ivo Dinov

Probabilities

- Computer programs provide lower-tail (or cumulative) probabilities of the form $\text{pr}(X \leq x)$
 - We give the program the x -value; it gives us the probability.
- Computer programs also provide inverse lower-tail probabilities (or quantiles)
 - We give the program the probability; it gives us the x -value.
- When calculating probabilities, we shade the desired area under the curve and then devise a way of obtaining it via lower-tail probabilities.

Slide 47 STAT 13, UCLA, Ivo Dinov

Standard Units

The z -score of a value a is

- the number of standard deviations a is away from the mean
- positive if a is above the mean and negative if a is below the mean.

The *standard Normal* distribution has $\mu = 0$ and $\sigma = 1$.

- We usually use Z to represent a random variable with a standard Normal distribution.

Slide 48 STAT 13, UCLA, Ivo Dinov

Ranges, extremes and z -scores

Central ranges:

- $P(-z \leq Z \leq z)$ is the same as the probability that a random observation from an arbitrary Normal distribution falls within z SD's either side of the mean.

Extremes:

- $P(Z \geq z)$ is the same as the probability that a random observation from an arbitrary Normal distribution falls more than z standard deviations above the mean.
- $P(Z \leq -z)$ is the same as the probability that a random observation from an arbitrary Normal distribution falls more than z standard deviations below the mean.

Slide 49 STAT 13, UCLA, Ivo Dinov

Combining Random Quantities

Variation and independence:

- No two animals, organisms, natural or man-made objects are ever identical.
- There is always variation. The only question is whether it is large enough to have a practical impact on what you are trying to achieve.
- Variation in component parts leads to even greater variation in the whole.

Slide 50 STAT 13, UCLA, Ivo Dinov

Independence

We model variables as being independent

- if we think they relate to physically independent processes
- and if we have no data that suggests they are related.

Both sums and differences of independent random variables are more variable than any of the component random variables

Slide 51 STAT 13, UCLA, Ivo Dinov

Formulas

- For a constant number a , $E(aX) = aE(X)$
and $SD(aX) = |a| SD(X)$.
- Means of sums and differences of random variables act in an obvious way
 - the mean of the sum is the sum of the means
 - the mean of the difference is the difference in the means
- For independent random variables, (cf. Pythagorean theorem),
 $SD(X_1 + X_2) = SD(X_1 - X_2) = \sqrt{SD(X_1)^2 + SD(X_2)^2}$
 $E(X_1 + X_2) = E(X_1) + E(X_2)$
 [ASIDE: Sums and differences of independent Normally distributed random variables are also Normally distributed]

Slide 52 STAT 13, UCLA, Ivo Dinov

Example

- Assumption: Crime rate for individuals is independent of family relations!
- Let X = RV representing the number of crimes an average individual commits in a 5 yr span.
- Let X_1 and X_2 be the crime rates of a husband and the wife in one family. What is the expected crime rate for this family given that $E(X) = 1.4$ and $SD(X) = 0.7$?
 $SD(X_1 + X_2) = SD(X_1 - X_2) = \sqrt{SD(X_1)^2 + SD(X_2)^2}$
 $E(X_1 + X_2) = E(X_1) + E(X_2)$

Slide 53 STAT 13, UCLA, Ivo Dinov

Example

- Assumption: Crime rate for individuals is independent of family relations!
- If $X_1 + X_2 = 3X - 1$, dependent case
- X_1 and X_2 are independent (family relation)
- Suppose $X_1 + X_2 = 5$, is this atypical?

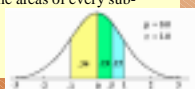
$$SD(X_1 + X_2) = SD(X_1 - X_2) = \sqrt{SD(X_1)^2 + SD(X_2)^2}$$

$$E(X_1 + X_2) = E(X_1) + E(X_2)$$

Slide 54 STAT 13, UCLA, Ivo Dinov

Areas under Standard Normal Curve – Normal Approximation

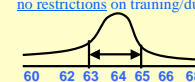
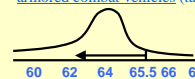
- Protocol:
 - Convert the interval (we need to assess the percentage of entries in) to **Standard units**. Actually convert the end points in Standard units.
 - In general, the transformation $X \rightarrow (X-\mu)/\sigma$, standardizes the observed value X , where μ and σ are the **average** and the **standard deviation** of the distribution X is drawn from.
 - Find the corresponding area under the normal curve (from tables or online databases);
 - Sketch the normal curve and shade the area of interest
 - Separate your area into individually computable sections
 - Check the Normal Table and extract the areas of every sub-section
 - Add/compute the areas of all sub-sections to get the total area.



Slide 55 STAT 13, UCLA, Ivo Dinov

Areas under Standard Normal Curve – Normal Approximation, Scottish Army Recruits

- The **mean height is 64 in** and the **standard deviation is 2 in**.
 - Only recruits shorter than 65.5 in will be trained for tank operation. **What percentage of the incoming recruits will be trained to operate armored combat vehicles (tanks)?**
 $X \rightarrow (X-64)/2$
 $65.5 \rightarrow (65.5-64)/2 = 1/4$
Percentage is 77.34%
 - Recruits within $1/2$ standard deviations of the mean will have **no restrictions on training/duties**. **About what percentage of the recruits will have no restrictions on training/duties?**
 $X \rightarrow (X-64)/2$
 $65 \rightarrow (65-64)/2 = 1/2$
 $63 \rightarrow (63-64)/2 = -1/2$
Percentage is 38.30%



Slide 56 STAT 13, UCLA, Ivo Dinov

Percentiles for Standard Normal Curve

- When the histogram of the observed process follows the normal curve Normal Tables (of any type, as described before) may be used to **estimate percentiles**. The **N-th percentile** of a distribution is **P** is **N%** of the population observations are less than or equal to **P**.
- Example, suppose the Math-part SAT scores of newly admitted freshmen at UCLA averaged 535 (out of [200:800]) and the SD was 100. Estimate the **95 percentile** for the score distribution.
- Solution:

Z	Area
1.65	90.11
1.70	91.09

Slide 57 STAT 13, UCLA, Ivo Dinov

Percentiles for Standard Normal Curve

- Example, suppose the Math-part SAT scores of newly admitted freshmen at UCLA averaged 535 (out of [200:800]) and the SD was 100. Estimate the **95 percentile** for the score distribution.
- Solution:

Z	Area
1.65	90.11
1.70	91.09
- $Z = 1.65$ (std. Units) $\rightarrow 700$ (data units), since

 $X \rightarrow (X - \mu) / \sigma$, converts data to standard units and

 $X \rightarrow \sigma X + \mu$, converts standard to data units!
- $\sigma = 100; \mu = 535; 100 \times 1.65 + 535 = 700$.

Slide 58 STAT 13, UCLA, Ivo Dinov

Summary

- The Standard Normal curve is symmetric w.r.t. the origin (0,0) and the total area under the curve is 100% (1 unit)
- Std units indicate how many SD's is a value below (-)/above (+) the mean
- Many histograms have roughly the shape of the normal curve (bell-shape)
- If a list of numbers follows the normal curve the percentage of entries falling within each interval is estimated by: 1. Converting the interval to StdUnits and, 2. Computing the corresponding area under the normal curve (Normal approximation)
- A histogram which follows the normal curve may be reconstructed just from (μ, σ^2) , **mean** and **variance**=std_dev²
- Any histogram can be summarized using percentiles
- $E(aX+b) = aE(X)+b$, $Var(aX+b) = a^2Var(X)$, where $E(Y)$ the mean of Y and $Var(Y)$ is the square of the StdDev(Y),

Slide 59 STAT 13, UCLA, Ivo Dinov

Example – work out in your notebooks

- Compute the chance a random observation from a distribution (symmetric, bell-shaped, unimodal) with $m=75$ and $SD=12$ falls within the range [53 : 71].

Check Work Should it be <50% or >50%?

Slide 60 STAT 13, UCLA, Ivo Dinov

Example – work out in your notebooks

Compute the chance a random observation from a distribution (symmetric, bellshaped & unimodal) with mean $\mu=75$ and $SD \sigma=12$, falls within the range [53 : 71].

- $(53-75)/12 = -11/6 = -1.83$ Std unit
- $(71-75)/12 = -0.333(3)$ Std units
- Area[53:71] =
- $(SN_area[-1.83 : 0]) - SN_area[-0.33 : 0])$
- $= (0.4664 - 0.1293) = 0.34$ (34%)
- Compute the 90th percentile for the same data:
- $b+a+c=100\%$ $a=40\%$
- $b=50\%$ $c=10\%$ } $Z=1.28$ SU
- 90% Percentile = $\sigma 1.3 + \mu = 12 \times 1.28 + 75 = 90.6$

Slide 61 STAT 13, UCLA, Ivo Dinov

General Normal Curve


- The general normal curve is defined by:
 - Where μ is the average of (the symmetric) normal curve, and σ is the standard deviation (spread of the distribution).
- Why worry about a standard and general normal curves?
- How to convert between the two curves?

$$y = \frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}}$$

Slide 62 STAT 13, UCLA, Ivo Dinov

Areas under Standard Normal Curve

- Many histograms are similar in shape to the **standard normal curve**. For example, persons height. The height of all incoming female army recruits is measured for custom training and assignment purposes (e.g., very tall people are inappropriate for constricted space positions, and very short people may be disadvantages in certain other situations). The mean height is computed to be 64 in and the standard deviation is 2 in. Only recruits shorter than 65.5 in will be trained for tank operation and recruits within 1/2 standard deviations of the mean will have no restrictions on duties.
 - What percentage of the incoming recruits will be trained to operate armored combat vehicles (tanks)?
 - About what percentage of the recruits will have no restrictions on training/duties?



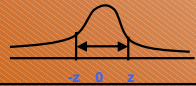
Slide 63 STAT 13, UCLA, Ivo Dinov

Standard Normal Curve – Table differences

- There are different tables and computer packages for representing the area under the **standard normal curve**. But the results are always **interchangeable**.

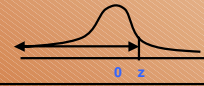
Area under Normal curve on $[-z : z]$

Z	Area
0.50	38.29
1.0	68.27




Area under Normal curve on $[-\infty : z]$

Z	Area
0.50	69.15
1.0	84.13



Area under Normal curve on $[z : \infty]$

Z	Area
0.50	30.85
1.0	15.87



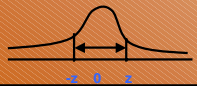
Slide 64 STAT 13, UCLA, Ivo Dinov

Standard Normal Curve – Table differences

- There are different tables and computer packages for representing the area under the **standard normal curve**. But the results are always **interchangeable**.

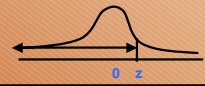
Area under Normal curve on $[-z : z]$

Z	Area
0.50	38.29
1.0	68.27




Area under Normal curve on $[-\infty : z]$

Z	Area
0.50	69.15
1.0	84.13



Area under Normal curve on $[0 : z]$

Z	Area
0.50	30.85
1.0	15.87



Slide 65 STAT 13, UCLA, Ivo Dinov